Sparse Channel Estimation for MIMO-OFDM Amplify-and-Forward Two-Way Relay Networks

Guan Gui, Abolfazl Mehbodniya and Fumiyuki Adachi
Department of Communications Engineering,
Graduate School of Engineering, Tohoku University
6-6-05 Aza-Aoba, Aramaki, Aoba-ku, Sendai, 980-8579 Japan
E-mail: {gui, mehbod}@mobile.ecei.tohoku.ac.jp, adachi@ ecei.tohoku.ac.jp

***Abstract*—Accurate channel impulse response (CIR) is required for coherent detection and it also helps to improve the quality of service in next-generation wireless communication systems. Linear channel estimation methods, e.g., least square (LS), have been proposed to estimate the CIR. However, these methods never take advantage of the channel sparsity and they also cause performance loss. In this paper, we propose a sparse channel estimation method for multi-input multi-output orthogonal frequency-division multiplexing (MIMO-OFDM) amplify and forward two-way relay networks (AF-TWRN), to exploit the sparse structure information in the CIR for each user. Sparse channel estimation problem is formulated as compressed sensing (CS) using sparse decomposition theory and the estimation process is implemented by LASSO algorithm. Computer simulation results are given to confirm the superiority of the proposed method over the LS-based channel estimation.**

Keywords—sparse Channel Estimation; MIMO-OFDM; AF-TWRN; compressed sensing (CS)

I. INTRODUCTION

Wireless communication technologies are developing rapidly due to the huge market promotion by the skyrocketing number of wireless users in last decades. So far, there has been three promising techniques for broadband wireless communications. The first technique is multiple antenna transmission. Multi-input multi-output (MIMO) is becoming one of the prevail techniques for enhancing the system capacity and combating multipath channel fading. The second technique is orthogonal frequency division multiplexing (OFDM) modulation which provides high spectral efficient and robustness to frequency-selective channel fading [1]. The third technique is two-way relay network (TWRN) that implements information exchange in two time slots. When comparing with four-time-slots traditional TWRN (see Fig. 1(a)) and three time slots physical-layer TWRN (see Fig. 1(b)) which achieve information exchange, two time slots TWRN (see Fig. 3(c)) can enhance system capacity by 66.7% and 100%, respectively. In addition, TWRN can also improve transmission range with limited transmit power [2]. Combining these three techniques together is a promising candidate technique for future advanced wireless communications. However, one of the key challenges is how to obtain accurate channel state information (CSI) which is needed for self-interference removal and coherent detection at each terminal.

The decode-and-forward (DF)-based TWRN system needs channel estimation at both terminal and relay station for coherent detection. Channel estimation and signal modulation will cost high computation burden on the relay. In addition, channel estimation techniques in DF MIMO-OFDM-TWRN [3] can be borrowed from point-to-point MIMO-OFDM systems [4–8]. Unlike the DF model, the obvious advantage of AF is that it can alleviate the computational burden on the relay, i.e., the relay amplifies and forwards the signals received from both terminals. Due to the signaling rule, only the cascaded channels are necessary for self-interference removal and coherent detection at each terminal. Estimating the cascaded channels at each terminal mitigates the quantization error, reduces the computational burden at the relay and avoids further channel distortion from noise [9]. Traditional linear channel estimation methods, e.g, LS [10], are proposed for MIMO-OFDM AF-TWRN . However, these methods do not take advantage of the inherent channel sparsity and hence cause performance loss.

In this paper, we study the channel estimation in a cyclic prefix (CP)-based MIMO-OFDM TWRN. Specifically, we propose a novel approach to exploit the channel sparsity. We formulate the problem using compressed sensing (CS). At each terminal, equivalent training signal is constructed to probe equivalent channel vector using least absolute shrinkage and selection operator (LASSO) [11]. The performance of proposed method is evaluated by computer simulations.

The remainder of this paper is organized as follows. A MIMO-OFDM AF-TWRN system model is described and problem formulation is given in Section II. In section III, the sparse channel estimation method is proposed and lower bound of estimation performance is derived. Computer simulation results are given in Section IV in order to evaluate and compare the performance of LS-based channel estimation method. Finally, we conclude the paper in Section V.

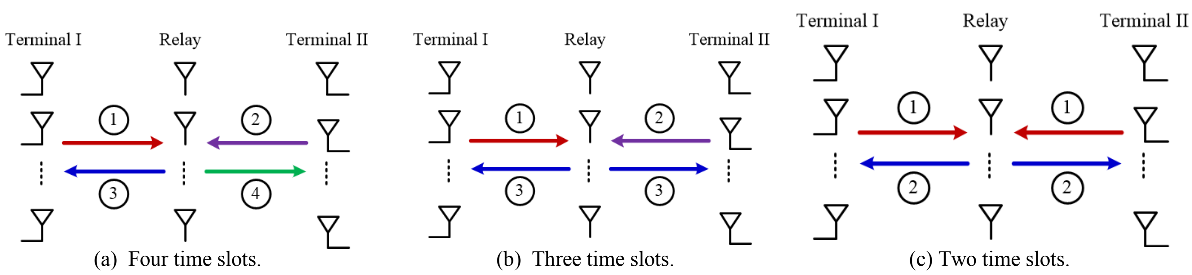
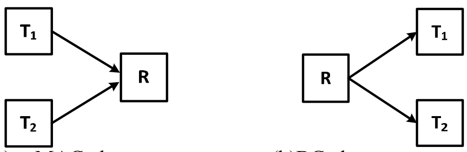


Fig. 1. Information exchange using different numbers of time slots in MIMO-OFDM AF-TWRN.

978-1-4673-6187-3/13/$31.00 ©2013 IEEE

II. SYSTEM MODEL AND PROBLEM FORMUALTION

As shown in Fig. 1(c), we consider a MIMO-OFDM AF-TWRN in which two-time-slot information exchange between terminal $\mathbf{T}_1$ and terminal $\mathbf{T}_2$ with the help of relay $\mathbf{R}$ is assumed. Both the terminals and the relay have N_t and N_r antennas ($N_r \geq N_t$), respectively. Assume that L -length channel vectors between the n_t -th antenna of terminals $\mathbf{T}_i$, $i = 1{,}2$ and the n_r-th antenna of relay $\mathbf{R}$ are denoted by $\mathbf{h}_{n_t n_r} = [h_{n_t n_r}(0), h_{n_t n_r}(1), \dots, h_{n_t n_r}(L-1)]^T$ and $\boldsymbol{g}_{n_t n_r} = [g_{n_t n_r}(0), g_{n_t n_r}(1), \dots, g_{n_t n_r}(L-1)]^T$, respectively. Each channel vector is supported by K nonzero taps and $K \ll L$. Suppose that each of the nonzero taps is modeled as a complex Gaussian random variable with zero mean and variance $\sigma_{h,l}^2$, and $\sigma_{g,l}^2, l = 0,1, \dots, L-1$. In addition, $\mathbf{h}_{n_t n_r}$ and $\boldsymbol{g}_{n_t n_r}$ are assumed to be invariant in the two time slots information exchange. At time t, suppose that OFDM signal vectors are transmitted from the n_t-th antenna of terminal $\mathbf{T}_i$, $i = 1, 2$, $\bar{\mathbf{s}}_{n_t} = [\bar{s}_{n_t}(0), \bar{s}_{n_t}(1), \dots, \bar{s}_{n_t}(N-1)]^T$ and $\bar{\mathbf{x}}_{n_t} = [\bar{x}_{n_t}(0), \bar{x}_{n_t}(1), \dots, x_{n_t}(N-1)]^T$, respectively, where N is the number of subcarriers and $n_r = 1{,}2{,}..{,}N_r$. At the same time, the transmit power is assumed $\mathrm{E}[\bar{\mathbf{s}}_{n_t}^H \bar{\mathbf{s}}_{n_t}] = P_1$ and $\mathrm{E}[\bar{\mathbf{x}}_{n_t}^H \bar{\mathbf{x}}_{n_t}] = P_2$, respectively.



(a) MAC phase. (b)BC phase.

Fig. 2. Information exchanges under TWNR.

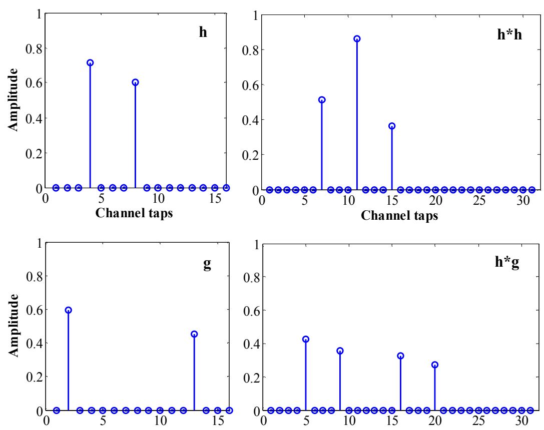


Fig. 3. Example of two individual channels and their cascaded ones.

A. MAC phase

In the multi-access (MAC) phase shown in Fig. 2(a), inverse discrete Fourier transform (IDFT) is computed for frequency-domain signal vectors $\bar{\mathbf{s}}_{n_t}$ and $\bar{\mathbf{x}}_{n_t}$. The resultant vectors, $\mathbf{s}_{n_t} = \mathbf{F}^H \bar{\mathbf{s}}_{n_t}$ and $\mathbf{x}_{n_t} = \mathbf{F}^H \bar{\mathbf{x}}_{n_t}$, are then cyclic prefix (CP) padded with CP length $L_{CP} \geq (L-1)$ to avoid inter-block interference (IBI). Here, $\mathbf{F}$ is an $N \times N$ discrete Fourier transform (DFT) matrix where entries $f_{mn} = 1/N\, e^{-j\pi mn/N}$, $m, n = 0{,}1, \dots, N$. After the CP removal, the received signal vector at the n_r-th antenna of $\mathbf{R}$ for $t = 1{,}2, \dots, T$ is written as

$$\mathbf{r}_{n_r} = \sum_{n_t}^{N_t} \mathbf{H}_{n_t n_r} \mathbf{s}_{n_t} + \sum_{n_t}^{N_t} \mathbf{G}_{n_t n_r} \mathbf{x}_{n_t} + \mathbf{z}_{n_r}, \tag{1}$$

for $n_r = 1{,}2{,}\dots, N_r$, where $\mathbf{H}_{n_t n_r}$ and $\mathbf{G}_{n_t n_r}$ are circultant matrices with the first columns equal to $[\mathbf{h}_{n_t n_r}^T, \mathbf{0}_{1\times(N-L)}]^T$, and $[\mathbf{g}_{n_t n_r}^T, \mathbf{0}_{1\times(N-L)}]^T$, respectively. The additive noise vector $\mathbf{z}_{n_r}$satisfies $\mathcal{CN}(\mathbf{0}_{N\times 1}, \sigma_n^2 \mathbf{I}_N)$. If we collect all received signal vectors $\mathbf{r}_{n_r}$ $n_r = 1{,}2, \dots, N_r$ at $\mathbf{R}$ to form a $N_r N$-length vector $\mathbf{r} = [\mathbf{r}_1^T, \dots, \mathbf{r}_{n_r}^T, \dots, \mathbf{r}_{N_r}^T]^T$, then the received signal at relay in the MAC phase is written as

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{G}\mathbf{x} + \mathbf{z}, \tag{2}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{21} & \dots & \mathbf{H}_{N_t 1} \\ \mathbf{H}_{12} & \mathbf{H}_{22} & \dots & \mathbf{H}_{N_t 2} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{H}_{1N_r} & \mathbf{H}_{2N_r} & \dots & \mathbf{H}_{N_t N_r} \end{bmatrix} \in \mathbb{C}^{N_r N \times N_t N}, \tag{3}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{21} & \dots & \mathbf{G}_{N_t 1} \\ \mathbf{G}_{12} & \mathbf{G}_{22} & \dots & \mathbf{G}_{N_t 2} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{G}_{1N_r} & \mathbf{G}_{2N_r} & \dots & \mathbf{G}_{N_t N_r} \end{bmatrix} \in \mathbb{C}^{N_r N \times N_t N}, \tag{4}$$

$$\mathbf{s} = [\mathbf{s}_1^T, \dots, \mathbf{s}_{n_t}^T, \dots, \mathbf{s}_{N_t}^T]^T \in \mathbb{C}^{N_t N \times 1}, \tag{5}$$

$$\mathbf{x} = [\mathbf{x}_1^T, \dots, \mathbf{x}_{n_t}^T, \dots, \mathbf{x}_{N_t}^T]^T \in \mathbb{C}^{N_t N \times 1}, \tag{6}$$

$$\mathbf{z} = [\mathbf{z}_1^T, \dots, \mathbf{z}_{n_r}^T, \dots, \mathbf{z}_{N_r}^T]^T \in \mathbb{C}^{N_r N \times 1}. \tag{7}$$

According to Eq. (2), the received signal vector $\mathbf{r}_{n_r}$ is amplified by a positive coefficients β which is given by

$$\beta = \sqrt{\frac{P_r}{N_t \sum_{l=0}^{L-1} \left(\sigma_{h,l}^2 P_1 + \sigma_{g,l}^2 P_2\right) + N_0}}, \tag{8}$$

where P_r is relay's amplify power which is given by $\mathrm{E}[\bar{\mathbf{r}}_{n_r}^H \bar{\mathbf{r}}_{n_r}] = P_r$.

B. BC phase

Because of system symmetry in TWRN, without loss of generality, we consider the broadcasting (BC) phase at $\mathbf{T}_1$, as shown in Fig.2(b). Let $\mathbf{y}_{n_t}$ denote the received signal vectors at the n_t-th antenna at time $(t+T)$. If we collect N_t received vectors $\mathbf{y}_{n_t}$ as $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_{N_t}^T]^T$, then received signal vector can be written as

$$\mathbf{y} = \beta \tilde{\mathbf{H}} \mathbf{H} \mathbf{s} + \beta \tilde{\mathbf{H}} \mathbf{G} \mathbf{x} + \beta \tilde{\mathbf{H}} \mathbf{z} + \mathbf{v}, \tag{9}$$

where

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \dots & \mathbf{H}_{1N_r} \\ \mathbf{H}_{21} & \mathbf{H}_{22} & \dots & \mathbf{H}_{2N_r} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{H}_{N_t 1} & \mathbf{H}_{N_t 2} & \dots & \mathbf{H}_{N_t N_r} \end{bmatrix} \in \mathbb{C}^{N_t N \times N_r N}, \tag{10}$$

$$\mathbf{v} = [\mathbf{v}_1^T, \mathbf{v}_2^T, \dots, \mathbf{v}_{N_t}^T]^T \in \mathbb{C}^{N_t N \times 1} \tag{11}$$

where $\mathbf{v}_{n_t}$ is a noise vector at the n_t -th antenna of $\mathbf{T}_1$, satisfying $\mathbf{v}_{n_t} \in \mathcal{CN}(\mathbf{0}_{N\times 1}, \sigma_n^2 \mathbf{I}_N)$. According to matrix theory [12], circulant matrices $\mathbf{H}_{n_t n_r}$ and $\mathbf{G}_{n_t n_r}$, $n_r = 1, 2, \dots, N_r$, $n_t = 1, 2, \dots, N_t$, can be decomposed as

$$\mathbf{H}_{n_t n_r} = \mathbf{F}^H \mathbf{\Lambda}_{n_t n_r} \mathbf{F}, \tag{12}$$

$$\mathbf{G}_{n_t n_r} = \mathbf{F}^H \mathbf{U}_{n_t n_r} \mathbf{F}, \tag{13}$$

where $(\cdot)^H$ denotes matrix Hermitian transition operation and above diagonal matrices are given by

$$\mathbf{\Lambda}_{n_t n_r} = \mathrm{diag}\{H_{n_t n_r}(0), \dots, H_{n_t n_r}(n), \dots, H_{n_t n_r}(N-1)\}, \tag{14}$$

$$\mathbf{U}_{n_t n_r} = \mathrm{diag}\{G_{n_t n_r}(0), \dots, G_{n_t n_r}(n), \dots, G_{n_t n_r}(N-1)\}. \tag{15}$$

Based on the above analysis, it is easily found that the n-th diagonal entries $H_{n_t n_r}(n)$ in Eq. (14) and $G_{n_t n_r}(n)$ in Eq. (15) are obtained by

$$H_{n_t n_r}(n) = \sum_{l=0}^{L-1} h_{n_t n_r}(n) e^{-j2\pi nl/N}, \tag{16}$$

$$G_{n_t n_r}(n) = \sum_{l=0}^{L-1} g_{n_t n_r}(n) e^{-j2\pi nl/N}. \tag{17}$$

Therefore, the product of $\beta \mathbf{H}_{n_t n_r} \mathbf{H}_{n_t' n_r'}$ and $\beta \mathbf{G}_{n_t n_r} \mathbf{H}_{n_t' n_r'}$ with respect to $n_t, n_t' = 1,2,\dots,N_t$ and $n_r = 1,2,\dots,N_r$ can also be written as

$$\beta \widetilde{\mathbf{H}}_{n_t n_r} \mathbf{H}_{n_t' n_r} = \mathbf{F}^H \beta \mathbf{\Lambda}_{n_t n_r} \mathbf{\Lambda}_{n_t' n_r} \mathbf{F}, \tag{18}$$

$$\beta \widetilde{\mathbf{H}}_{n_t n_r} \mathbf{G}_{n_t n_r} = \mathbf{F}^H \beta \mathbf{\Lambda}_{n_t n_r} \mathbf{U}_{n_t n_r} \mathbf{F}. \tag{19}$$

Hence, both $\beta \widetilde{\mathbf{H}}_{n_t n_r} \mathbf{H}_{n_t' n_r}$ and $\beta \widetilde{\mathbf{H}}_{n_t n_r} \mathbf{G}_{n_t n_r}$ are circulant matrices where their first columns are given by $[\beta(\mathbf{h}_{n_t n_r} * \mathbf{h}_{n_t' n_r})^T \ \mathbf{0}_{1\times(N-2L+1)}]^T$ and $[\beta(\mathbf{h}_{n_t n_r} * \boldsymbol{g}_{n_t n_r})^T \ \mathbf{0}_{1\times(N-2L+1)}]^T$, respectively, where '*' denotes convolution operator between two channel vectors. Based on this observation, when the n_1-th row vector of $\beta \widetilde{\mathbf{H}}$ multiplies with the n_2-th column vector of $\mathbf{H}$, $n_t, n_t' = 1,2,\dots,N_t$, we can obtain an equivalent ($2L$-1)-length cascaded channel vector $\mathbf{q}_{n_t' n_t} \triangleq [q_{n_t' n_t}(0), \dots, q_{n_t' n_t}(l), \dots, q_{n_t' n_t}(2L-2)]^T$ which is given by

$$\mathbf{q}_{n_t' n_t} \triangleq \beta \sum_{n_r=1}^{N_r} \mathbf{h}_{n_t n_r} * \mathbf{h}_{n_t' n_r}. \tag{20}$$

Because of the symmetry of two MIMO channel matrices, we can easily find their symmetry relationship, that is, $\mathbf{q}_{n_t' n_t} = \mathbf{q}_{n_t n_t'}$. Hence, the product $\beta \widetilde{\mathbf{H}}\mathbf{H}$ is equivalent to provide $(N_t^2 + N_t)/2$ independent $(2L-1)$-length composite channel vectors $\mathbf{q}_{n_t' n_t}$ with $n_t, n_t' = 1, 2, \dots, N_t$. Note that $(N_t^2 + N_t)/2 < N_t^2$ if $N_t > 1$. By virtual of the duplication matrix property [13] on sparse channel estimation, the complexity reduces, especially in the case of a relatively large scale communication system. Mathematically speaking, the computational complexity reduces to $O((N_t^2 + N_t)/2)$ rather than $O(N_t^2)$, where $O(\cdot)$ denotes the calculation metric of complexity. Due to independence between the two MIMO channel matrices $\widetilde{\mathbf{H}}$ and $\mathbf{G}$, hence, $\beta \widetilde{\mathbf{H}}\mathbf{G}$ is equivalent to generate N_t^2 independent ($2L$-1)-length cascaded channel vectors $\mathbf{p}_{n_t n_t'} \triangleq [p_{n_t n_t'}(0), \dots, p_{n_t n_t'}(l), \dots, p_{n_t n_t'}(2L-2)]^T$ with respect to $n_t, n_t' = 1, 2, \dots, N_t$, where

$$\mathbf{p}_{n_t n_t'} \triangleq \beta \sum_{n_r=1}^{N_r} \mathbf{h}_{n_t n_r} * \boldsymbol{g}_{n_t' n_r}. \tag{21}$$

If we define $\tilde{\mathbf{F}} = \mathbf{I}_{N_t} \otimes \mathbf{F} \in \mathbb{C}^{N_t N \times N_r N}$, where '$\otimes$' denotes Kronecker product and $\mathbf{I}_{N_t}$ denotes an $N_t \times N_t$ identity matrix, the received signal $\mathbf{y}$ in Eq. (9) is transformed to frequency-domain using DFT matrix $\tilde{\mathbf{F}}$, then, we have

$$\bar{\mathbf{y}} = \tilde{\mathbf{F}} \beta \widetilde{\mathbf{H}} \mathbf{H} \tilde{\mathbf{F}}^H \bar{\mathbf{s}} + \tilde{\mathbf{F}} \beta \widetilde{\mathbf{H}} \mathbf{G} \tilde{\mathbf{F}}^H \bar{\mathbf{x}} + \bar{\mathbf{v}}, \tag{22}$$

where $\bar{\mathbf{v}} = \tilde{\mathbf{F}} \beta \widetilde{\mathbf{H}} \mathbf{z} + \tilde{\mathbf{F}} \mathbf{v}$ denotes composite noise vector at the $\mathbf{T}_1$. According to Eq. (18) and (19), $\tilde{\mathbf{F}} \beta \widetilde{\mathbf{H}} \mathbf{H} \tilde{\mathbf{F}}^H$ and $\tilde{\mathbf{F}} \beta \widetilde{\mathbf{H}} \mathbf{G} \tilde{\mathbf{F}}^H$ can be given in Eq. (23) and (24),

$$\tilde{\mathbf{F}} \beta \widetilde{\mathbf{H}} \mathbf{H} \tilde{\mathbf{F}}^H = \begin{bmatrix} \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{1n_r} \mathbf{\Lambda}_{1n_r} & \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{1n_r} \mathbf{\Lambda}_{2n_r} & \cdots & \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{1n_r} \mathbf{\Lambda}_{N_t n_r} \\ \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{2n_r} \mathbf{\Lambda}_{1n_r} & \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{2n_r} \mathbf{\Lambda}_{2n_r} & \cdots & \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{2n_r} \mathbf{\Lambda}_{N_t n_r} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{N_t n_r} \mathbf{\Lambda}_{1n_r} & \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{N_t n_r} \mathbf{\Lambda}_{2n_r} & \cdots & \sum_{n_r=1}^{N_r} \beta \mathbf{\Lambda}_{N_t n_r} \mathbf{\Lambda}_{N_t n_r} \end{bmatrix}, \tag{23}$$

$$\tilde{\mathbf{F}} \beta \widetilde{\mathbf{H}} \mathbf{G} \tilde{\mathbf{F}}^H = \begin{bmatrix} \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{1n_r} \mathbf{U}_{1n_r} & \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{1n_r} \mathbf{U}_{2n_r} & \cdots & \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{1n_r} \mathbf{U}_{N_t n_r} \\ \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{2n_r} \mathbf{U}_{1n_r} & \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{2n_r} \mathbf{U}_{2n_r} & \cdots & \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{2n_r} \mathbf{U}_{N_t n_r} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{N_t n_r} \mathbf{U}_{1n_r} & \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{N_t n_r} \mathbf{U}_{2n_r} & \cdots & \sum_{n_r=1}^{N_r} \boldsymbol{\beta} \mathbf{\Lambda}_{N_t n_r} \mathbf{\Lambda} \mathbf{U}_{N_t n_r} \end{bmatrix}, \tag{24}$$

respectively. If we define both $\mathbf{S}_i = \mathrm{diag}(\bar{\mathbf{s}}_i)$ and $\mathbf{X}_i = \mathrm{diag}(\bar{\mathbf{x}}_i)$ as $N \times N$ diagonal matrices, and collect all cascaded channel vectors as $\mathbf{q} \triangleq [\mathbf{q}_{11}^T, \dots, \mathbf{q}_{1N_t}^T, \mathbf{q}_{22}^T, \dots, \mathbf{q}_{2N_t}^T, \dots, \mathbf{q}_{N_t^2}^T]^T$ and $\mathbf{p} \triangleq [\mathbf{p}_{11}^T, \dots, \mathbf{p}_{N_t 1}^T, \dots, \mathbf{p}_{1N_t}^T, \dots, \mathbf{p}_{N_t^2}^T]^T$, then two equivalent training signal matrices can be written in Eq. (25) and (26) respectively, where $\mathbf{F}_{2L-1}$ is partial DFT matrix by extracting the first $(2L-1)$-columns of $\mathbf{F}$.

Then the received signal model in Eq. (22) can be reformulated as

$$\bar{\mathbf{y}} = \mathbf{S}\mathbf{q} + \mathbf{X}\mathbf{p} + \bar{\mathbf{v}} = \mathbf{D}\mathbf{b} + \bar{\mathbf{v}}, \tag{27}$$

where $\mathbf{D} = [\mathbf{S}, \mathbf{X}]$ denotes the equivalent training matrix which combines two training signal matrices $\mathbf{S}$ of $N_t N \times (2L-1)N_t(N_t+1)/2$ sizes and $\mathbf{X}$ of $N_t N \times (2L-1)N_t^2$ sizes, $\mathbf{b} = [\mathbf{q}^T, \mathbf{p}^T]^T$ denotes the overall channel vector including $\mathbf{q}$ and $\mathbf{p}$. At the receive side of $\mathbf{T}_1$, channel estimator $\mathbf{q}$ is used to remove self-data interference and channel estimator $\mathbf{p}$ is applied to extract other users' data information at $\mathbf{T}_1$.

According to the system model in Eq. (27), it is easily understood that the main objective of this paper is to estimate the overall channel vector $\mathbf{b}$ using the composite training signal matrix $\mathbf{D}$. With respect to Eq. (27), LS based channel estimator $\mathbf{b}_{LS}$ can be computed by

$$\mathbf{b}_{ls} = (\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H\bar{\mathbf{y}} = \mathbf{b} + (\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H\bar{\mathbf{v}}. \tag{28}$$

Since the noise variance of $\bar{\mathbf{v}}$ is given by

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1\mathbf{F}_{2L-1} & \mathbf{S}_2\mathbf{F}_{2L-1} & \mathbf{S}_3\mathbf{F}_{2L-1} & \cdots & \mathbf{S}_{N_t}\mathbf{F}_{2L-1} & \mathbf{0}_{N\times(2L-1)} & \mathbf{0}_{N\times(2L-1)} & \mathbf{0}_{N\times(2L-1)} \\ \mathbf{0}_{N\times(2L-1)} & \mathbf{S}_1\mathbf{F}_{2L-1} & \mathbf{S}_2\mathbf{F}_{2L-1} & \cdots & \mathbf{S}_{N_t-1}\mathbf{F}_{2L-1} & \mathbf{S}_{N_t}\mathbf{F}_{2L-1} & \mathbf{0}_{N\times(2L-1)} & \vdots \\ \vdots & \mathbf{0}_{N\times(2L-1)} & \ddots & \ddots & \vdots & \ddots & \ddots & \mathbf{0}_{N\times(2L-1)} \\ \mathbf{0}_{N\times(2L-1)} & \cdots & \mathbf{0}_{N\times(2L-1)} & \mathbf{S}_1\mathbf{F}_{2L-1} & \mathbf{S}_2\mathbf{F}_{2L-1} & \cdots & \mathbf{S}_{N_t-1}\mathbf{F}_{2L-1} & \mathbf{S}_{N_t}\mathbf{F}_{2L-1} \end{bmatrix}, \tag{25}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1\mathbf{F}_{2L-1} & \mathbf{0}_{N\times(2L-1)} & \mathbf{0}_{N\times(2L-1)} & \mathbf{0}_{N\times(2L-1)} & \cdots & \mathbf{X}_{N_t}\mathbf{F}_{2L-1} & \mathbf{0}_{N\times(2L-1)} & \mathbf{0}_{N\times(2L-1)} & \mathbf{0}_{N\times(2L-1)} \\ \mathbf{0}_{N\times(2L-1)} & \mathbf{X}_1\mathbf{F}_{2L-1} & \mathbf{0}_{N\times(2L-1)} & \vdots & \cdots & \mathbf{0}_{N\times(2L-1)} & \mathbf{X}_{N_t}\mathbf{F}_{2L-1} & \vdots & \vdots \\ \vdots & \mathbf{0}_{N\times(2L-1)} & \ddots & \mathbf{0}_{N\times(2L-1)} & \vdots & \ddots & \mathbf{0}_{N\times(2L-1)} & \ddots & \mathbf{0}_{N\times(2L-1)} \\ \mathbf{0}_{N\times(2L-1)} & \cdots & \mathbf{0}_{N\times(2L-1)} & \mathbf{X}_1\mathbf{F}_{2L-1} & \cdots & \mathbf{0}_{N\times(2L-1)} & \cdots & \mathbf{0}_{N\times(2L-1)} & \mathbf{X}_{N_t}\mathbf{F}_{2L-1} \end{bmatrix}, \tag{26}$$

$$\mathrm{E}\{\bar{\mathbf{v}}^H\bar{\mathbf{v}}\} = N_0\left(\beta^2 N_r \sum_{l=0}^{L-1}\sigma_{h,l}^2 + 1\right), \quad (29)$$

then the average MSE of LS channel estimator $\mathbf{b}_{LS}$ is given by

$$\mathrm{MSE}\{\mathbf{b}_{ls}\} = N_0\left(\beta^2 N_r \sum_{l=0}^{L-1}\sigma_{h,l}^2 + 1\right)\mathrm{Trace}\{(\mathbf{D}^H\mathbf{D})^{-1}\}. \quad (30)$$

It is known that the training matrix $\mathbf{D}$ has $N_t\,(3N_t+1)(2L-1)/2$ columns that are normalized in a way such that $\|\mathbf{D}\|_F^2 = N_t\,(3N_t+1)(2L-1)/2$, where $\|\cdot\|_F$ denotes the Frobenius norm. Optimal training design for LS-based channel estimation method is subject to $\mathbf{D}^H\mathbf{D} = \mathbf{I}_{N_t(3N_t+1)(2L-1)/2}$. Hence, we have

$$\mathrm{Trace}(\mathbf{D}^H\mathbf{D}) = \|\mathbf{D}\|_F^2 = N_t\,(3N_t+1)(2L-1)/2, \quad (31)$$

where $\mathrm{Trace}(\mathbf{A})$ is defined to be the sum of the elements on the main diagonal of matrix $\mathbf{A}$. According to arithmetic–harmonic means inequality, lower bound for the LS channel estimation error can be derived as

$$\mathrm{MSE}\{\mathbf{b}_{ls}\} \geq \frac{N_0\left(\beta^2 N_r \sum_{l=0}^{L-1}\sigma_{h,l}^2 + 1\right)\left(N_t\,(3N_t+1)(2L-1)/2\right)^2}{\mathrm{Trace}\{\mathbf{D}^H\mathbf{D}\}}$$

$$= N_0\left(\beta^2 N_r \sum_{l=0}^{L-1}\sigma_{h,l}^2 + 1\right)(3N_t+1)(2L-1)/2. \quad (32)$$

From the derivation in Eq. (32), the lower bound of LS can be written as $\mathrm{MSE}\{\mathbf{b}_{LS}\} \sim \boldsymbol{O}(N_0, \sum_{l=0}^{L-1}\sigma_{h,l}^2, \beta, N_r, N_t, L)$. Generally, linear channel estimation methods, e.g., LS, emphasize on optimal training design to improve the estimation performance while neglect the inherent sparsity of channel.

III. SPARSE CHANNEL ESTIMATION

According to the CS [14], [15], accurate sparse channel estimation requires that training signal matrix $\boldsymbol{D}$ be satisfied restricted isometric property (RIP) [16] in high probability. Hence, according to the system model in Eq. (27), optimal sparse channel estimator $\mathbf{b}_{opt}$ can be given by

$$\mathbf{b}_{opt} = \mathrm{argmin}_{\mathbf{b}}\left\{\tfrac{1}{2}\|\mathbf{D}\mathbf{b} - \bar{\mathbf{y}}\|_2^2 + \lambda\|\mathbf{b}\|_0\right\}, \quad (33)$$

where $\|\mathbf{b}\|_2$ denotes the Euclidean norm which is given by $\|\mathbf{b}\|_2^2 = \sum_i |b_i|^2$, $\|\mathbf{b}\|_0$ denotes the zero-norm operator which counts their nonzero taps and λ is the regularization parameter which trades off the estimation error and sparseness of the channel. Assume the position sets of all channel taps of $\mathbf{b}$ is $\boldsymbol{\Omega}$ and its nonzero taps set is Γ. The number of nonzero taps of $\mathbf{b}$ is T, then the lower bound of sparse channel estimator is derived by

$$\mathrm{MSE}\{\mathbf{b}_{opt}\} = N_0\left(\beta^2 N_r \sum_{l=0}^{L-1}\sigma_{h,l}^2 + 1\right)\mathrm{Trace}\{(\mathbf{D}_\Gamma^H\mathbf{D}_\Gamma)^{-1}\}$$

$$\geq \frac{N_0\left(\beta^2 N_r \sum_{l=0}^{L-1}\sigma_{h,l}^2 + 1\right)T}{\mathrm{Trace}\{\mathbf{D}^H\mathbf{D}\}}$$

$$= N_0\left(\beta^2 N_r \sum_{l=0}^{L-1}\sigma_{h,l}^2 + 1\right)T, \quad (34)$$

where $\mathrm{Trace}\{\mathbf{D}_\Gamma^H\mathbf{D}_\Gamma\} = \mathbf{I}_T$ denotes the optimal signal training for sparse channel estimation. Comparing Eq. (34) with Eq. (32), we can find that the lower bound of optimal channel estimator depends on T rather than overall channel length $N_t(3N_t+1)(2L-1)/2$ of $\mathbf{b}$. If we can estimate positions of nonzero taps of $\mathbf{b}$, then sparse channel estimation performance could be improved. Since solving the optimal sparse channel estimation in Eq. (33) is NP hard problem [15]. Hence, it is necessary to develop alternative suboptimal sparse channel estimation method.

In this paper, we propose a sparse channel estimation method for MIMO-OFDM AF-TWRN and it is implemented by LASSO algorithm [11]. Given an equivalent training matrix $\boldsymbol{D}$ and a received signal vector $\bar{\mathbf{y}}$, LASSO based sparse channel estimator $\mathbf{b}_{CS}$ can be obtained

$$\mathbf{b}_{cs} = \mathrm{argmin}_{\mathbf{b}}\left\{\tfrac{1}{2}\|\mathbf{D}\mathbf{b} - \bar{\mathbf{y}}\|_2^2 + \lambda\|\mathbf{b}\|_1\right\}, \quad (35)$$

where $\|\mathbf{b}\|_1$ denotes L_1-norm which is given by $\|\mathbf{b}\|_1 = \sum_i |b_i|$. In a practical system, accurate number of nonzero channel taps is unknown. Hence, to obtain an accurate sparse channel estimation, effective training signal design is required. In accordance with the CS [14], [15], two training design methods, i.e., random Gaussian and random binary, are considered for computer simulation to evaluate our method.

IV. COMPUTER SIMULATION

In this section, we present the simulation results to evaluate sparse channel estimation method in MIMO-OFDM AF-TWRN. Here we compare the performance of the proposed estimator with LS-based channel estimator and adopt 100 independent Monte-Carlo runs for averaging. The number (N_t, N_r) of transmitter/relay pairs are considered three cases: $(2{,}2)$, $(2{,}4)$ and $(4{,}2)$. All of the channel vectors have the same length $L = 16$ and $K = 1,2,3,4,5,6$, and its positions of nonzero channel taps are randomly generated. Training signal length of each antenna is set as $N = 32$ to ensure $N \geq 2L - 1$. Transmit power is set as $P_1 = P_2 = P$ and relay power is allocated as $P_r = 2P$. The signal to noise ratio (SNR) is defined as $10\log(P_r/\sigma_n^2)$at relay and $10\log(P_i/\sigma_n^2)$, $i = 1,2$ at transmitter, respectively.

Random Gaussian training is considered in Figs. 4 and 5, and random binary training is considered in Figs. 6 and 7. From the four figures, we can find that the proposed sparse channel estimator is better than the LS one. In addition, the four figures show that LS channel estimator depends on channel length while the proposed one relies on nonzero number K of channel. Note that the lower bound is given by ideal LS channel estimator which knows the positions of nonzero taps of the channel. In four experiments, the proposed sparse method works well on different number of nonzero taps of channel. However, for sparser channel estimation, more sparsity can be exploited. In other words, much better performance can be achieved. For example assuming $K = 1$, the proposed sparse channel estimator approaches to lower bound. On the contrary, channel is approximate sparse, e.g., $K = 6$, the performance advantage of the proposed method is no longer obvious. When $K = L = 16$, the proposed sparse channel estimator reduces to LS one. Because single channel vector between each pair of antennas is not exactly sparse, it will incur more number of nonzero taps in the cascaded channel. Hence, the proposed method can only work well in very sparse channel and when their cascaded channel is also satisfying sparse.

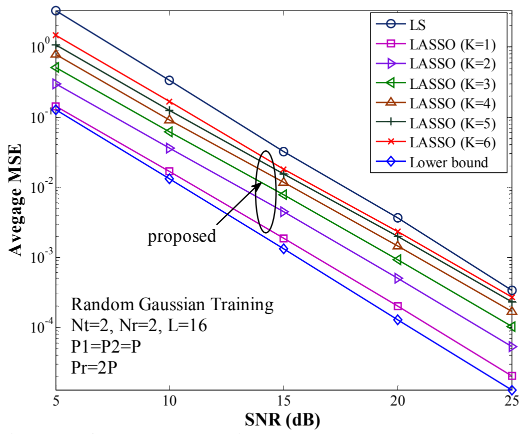


Fig. 4. Performance comparison versus SNR.

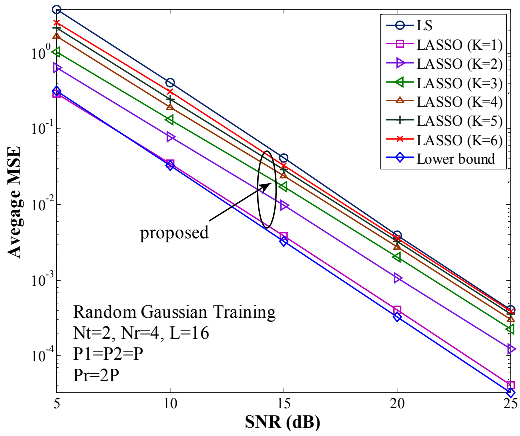


Fig. 5. Performance comparison versus SNR.

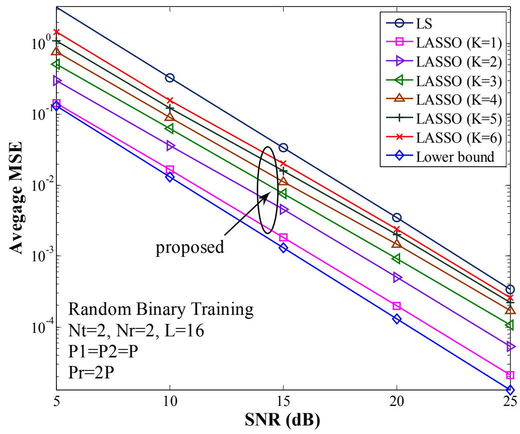


Fig.6. Performance comparison versus SNR.

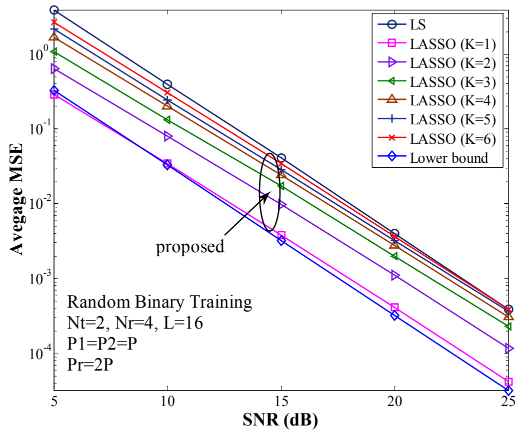


Fig. 7. Performance comparison versus SNR.

V. CONCLUSION

In this paper, we proposed a sparse channel estimation method which can exploit the extra knowledge of sparse structure as for prior information and hence it can increase the spectral efficiency or enhance the estimation performance when compared with traditional methods. Computer simulation results showed the performance advantages of our proposed method over LS based on the MSE standard.

REFERENCES

[1] D. Raychaudhuri and N. B. Mandayam, "Frontiers of wireless and mobile communications," *Proc. of the IEEE*, vol. 100, no. 4, pp. 824–840, Apr. 2012.

[2] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference:analog network coding," Computer Science and Artificial Intelligence, Laboratory Technical Report, MIT-CSAIL-TR-2007-012, Feb. 23, 2007

[3] M. Eslamifar, W. H. Chin, C. Yuen, and Y. L. Guan, "Performance analysis of two-step bi-directional relaying with multiple antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4237–4242, Dec. 2012.

[4] J. D. Gibson and R. A. Iltis, "Channel estimation and data detection for MIMO-OFDM systems," in *IEEE GLOBECOM, San Francisco, USA, Dec. 1-5*, 2003, no. l, pp. 581–585.

[5] H. Tuan, H. Kha, and H. Nguyen, "Optimized training sequences for spatially correlated MIMO-OFDM," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2768–2778, Sept. 2010.

[6] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.

[7] T. Pham, H. K. Garg, Y. Liang, and A. Nallanathan, "Joint channel estimation and data detection for MIMO-OFDM two-way relay networks," in *IEEE GLOBECOM*, Miami, USA, Dec. 6-10, 2010, pp.1–5.

[8] C. Shen, S. Member, M. P. Fitz, S. Member, and A. The, "MIMO-OFDM beamforming for improved channel estimation," *IEEE J. Sel. Areas in Commun.*, vol. 26, no. 6, pp. 948–959, Aug. 2008.

[9] F. Gao, R. Zhang, and Y. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.

[10] T.-H. Pham and Y.-C. Liang, "Channel estimation and training design for MIMO-OFDM two-way relay systems," in IEEE ICC, Ottawa, ON, June 10-15, 2012, pp. 3703–3707.

[11] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. Royal Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[12] R. M. Gray, "Toeplitz and Circulant Matrices: A review," *Foundations and Trends in Commun. and Inf. Theory*, vol. 3, no. 2, pp. 155–239, Jan. 2006.

[13] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley., vol. 31, no. 4. John Wiley & Sons, Ltd, 1989.

[14] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles□: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[15] D. L. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[16] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9–10, pp. 589–592, May 2008